



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **09251368 A**(43) Date of publication of application: **22.09.97**

(51) Int. Cl. **G06F 3/14**  
**G06F 3/02**  
**G06F 9/06**  
**G06F 9/45**

(21) Application number: **08060593**(22) Date of filing: **18.03.96**(71) Applicant: **HITACHI LTD**

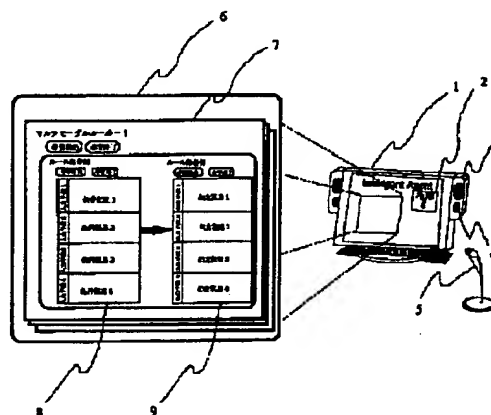
(72) Inventor: **TANO SHUNICHI**  
**SAKAO HIDEKI**  
**NANBA YASU HARU**  
**TOMITA TAMINORI**  
**AOSHIMA HIROKAZU**

**(54) CONSTRUCTION SUPPORTING SYSTEM FOR INTELLIGENT AGENT****(57) Abstract:**

**PROBLEM TO BE SOLVED:** To provide a function for freely constructing an agent responding by using plural media such as still images, moving images, voice and effective sound, etc., by controlling a computer and an application software, etc., instead of a user based on an instruction using the plural media such as a keyboard, a mouse, a pen, the voice and gestures, etc., even by the user incapable of performing programming.

**SOLUTION:** By writing descriptions for input contents and input timings in the plural media such as keyboard input, mouse input, pen input, voice input and gesture input, etc., in a rule condition part 8 and writing the output information descriptions, execution command descriptions and the output timing descriptions of text display output, still image output, moving image output, voice output and effective sound output, etc., in a rule conclusion part 9, the operation of the agent is described.

COPYRIGHT: (C)1997,JPO



(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/14	3 4 0		G 0 6 F 3/14	3 4 0 A
3/02	3 8 0		3/02	3 8 0 B
9/06	5 3 0		9/06	5 3 0 P
9/45			9/44	3 2 0 A

審査請求 未請求 請求項の数 6 O L (全 9 頁)

(21) 出願番号 特願平8-60593

(22) 出願日 平成8年(1996)3月18日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 田野 俊一

神奈川県川崎市麻生区王禅寺1099番地株式

会社日立製作所システム開発研究所内

(72) 発明者 坂尾 秀樹

神奈川県川崎市麻生区王禅寺1099番地株式

会社日立製作所システム開発研究所内

(72) 発明者 藤波 康晴

神奈川県川崎市麻生区王禅寺1099番地株式

会社日立製作所システム開発研究所内

(74) 代理人 弁理士 小川 勝男

最終頁に続く

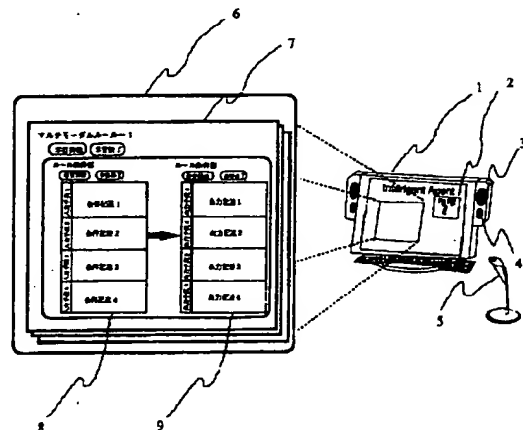
(54) 【発明の名称】 知的エージェントの構築支援システム

## (57) 【要約】

【課題】プログラミングを出来ないユーザでも、キーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答するエージェントを自由自在に構築するための機能を提供することにある。

【解決手段】ルール条件部8に、キーボード入力、マウス入力、ペン入力、音声入力、身振り入力など複数のメディアにおける入力内容および入力タイミングに関する記述を書き、ルール結論部9に、テキスト表示出力、静止画出力、動画出力、音声出力、効果音出力等の出力情報記述、実行コマンド記述、それらの出力タイミング記述を書くことにより、エージェントの動作を記述出来るようにした。

図1



## 【特許請求の範囲】

【請求項1】 キーボード、マウス、ペン、音声、身振りなど複数のメディアを用いたユーザの指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答するエージェントを有するコンピュータシステムにおいて、

キーボード入力、マウス入力、ペン入力、音声入力、身振り入力など複数のメディアにおける入力内容および入力タイミングに関する記述からなるマルチモーダル条件記述と、その条件記述が満たされたとき実行あるいは応答すべき内容を表わす、テキスト表示出力、静止画出力、動画出力、音声出力、効果音出力等の出力情報記述、実行コマンド記述、それらの出力タイミング記述等から構成されるマルチモーダル実行部記述、の2つの組で構成されるマルチモーダルルールの集合により、エージェントの動作が規定されていることを特徴とする知的エージェントの構築支援システム。

【請求項2】 請求項1記載のシステムにおいて、マルチモーダル条件記述は、共通の時間軸上に、音声やキーボード入力等の各入力モードごとに入力内容指定が可能であることを特徴とする知的エージェントの構築支援システム。

【請求項3】 請求項1記載のシステムにおいて、マルチモーダル実行記述は、共通の時間軸上に、音声や画面出力等の各出力モードごとに出力内容指定が可能であることを特徴とする知的エージェントの構築支援システム。

【請求項4】 請求項1記載のシステムにおいて、マルチモーダル条件記述およびマルチモーダル出力記述にコンピュータ画面の状態の指定を含み、マルチモーダル条件記述におけるコンピュータ画面の記述は、そのコンピュータ画面が実際にコンピュータの出力画面に現われた時点に真となる条件として、マルチモーダル出力記述に現われるコンピュータ画面の記述は、そのコンピュータ画面となるようにコンピュータを制御したり、ソフトウェアアプリケーションを操作するコマンドの実行として、解釈されることを特徴とする知的エージェントの構築支援システム。

【請求項5】 請求項1記載のシステムにおいて、ユーザが複数の正しい例や誤った例をマルチモーダルに与えることによりマルチモーダル条件部記述やマルチモーダル実行部記述を自動生成することを特徴とする知的エージェントの構築支援システム。

【請求項6】 請求項1記載のシステムにおいて、ユーザの操作履歴をキーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた入力情報や、静止画、動画、音声、効果音等の出力情報、ユーザの操作等からなるマルチモーダル操作ログを記憶し、解析することにより、マルチモーダルルールを自動生成することを特徴とする知的エージェントの構築支援システム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、パーソナルコンピュータや携帯情報機器等のユーザインタフェース技術に関し、より詳細には、キーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答するエージェントの構築技術に関する。

## 【0002】

【従来の技術】 現在、パーソナルコンピュータや携帯情報機器等では、キーボードやマウスを用いた入力が主流である。しかし、キーボードやマウスの入力は繁雑であるため、ユーザに代わって、コンピュータやアプリケーションの操作を代行してくれるエージェントと呼ばれるソフトウェアが注目を浴びている。

【0003】 例えば、文献(a)「P. Mayes: Agents that Reduce Work and Information Overload, Communications of the ACM, Vol, 37, No. 7, pp. 30-40, 1994」では、大量に送付される電子メールを読むべきか、捨てるべきかをアドバイスするエージェントが記述されている。このエージェントはユーザの肩越しにユーザのキーボードとマウスの操作を監視しており、ユーザの操作の癖を学習する。

【0004】 また、文献(b)「D. Smith, A. Cypther and J. Spohrer: Kid Sim: Programming Agents Without a Programming Language, Communications of the ACM, Vol, 37, No. 7, pp. 54-67, 1994」では、グラフィカルルールを用いて、子供でもプログラミング出来る手法を提案しており、これを応用すればエージェントの構築に使えと主張している。

## 【0005】

【発明が解決しようとする課題】 しかし、上記技術は、キーボード、マウスに特化されており、キーボード、マウスを含む、ペン、音声、身振りなど複数のメディアを用いた指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答するエージェントへの適用は困難であった。

【0006】 本発明の目的は、プログラミングができないユーザでも、キーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答するエージェントを自由自在に構築するための機能を提供することにある。

## 【0007】

【課題を解決するための手段】 上記目的を達成するた

め、(1)プログラミングができないユーザでも、マルチモーダル入力に反応し、マルチモーダル出力で応答するエージェントを容易に作ることを可能とするために、ルール条件部に、キーボード入力、マウス入力、ペン入力、音声入力、身振り入力など複数のメディアにおける入力内容および入力タイミングに関する記述を書きルール結論部に、テキスト表示出力、静止画出力、動画出力、音声出力、効果音出力等の出力情報記述、実行コマンド記述、それらの出力タイミング記述を書くことにより、エージェントの動作を記述できるようにしたこと、

(2)マルチモーダルなルール条件記述を可能とするために、共通の時間軸上に、音声やキーボード入力等の各入力モードごとに入力内容指定を可能としたこと、

(3)マルチモーダルなルール結論部記述を可能とするために、共通の時間軸上に、音声や画面出力等の各出力モードごとに出力内容指定を可能としたこと、(4)ユーザはコンピュータの出力画面を見て操作を決めており、さらに、ユーザの操作結果はコンピュータ出力画面の状態変化として記述できるため、マルチモーダル条件

記述およびマルチモーダル出力記述にコンピュータ画面の状態の指定を可能とし、マルチモーダル条件記述におけるコンピュータ画面の記述は、そのコンピュータ画面が実際にコンピュータの出力画面に現われた時点に真となる条件として、マルチモーダル出力記述に現われるコンピュータ画面の記述は、そのコンピュータ画面となるようにコンピュータを制御したり、アプリケーションソフトウェアを操作するコマンドの実行として、解釈すること、(5)また、上記のようにマルチモーダルルールをユーザが明示的に記述しなくても、ルール条件部やルール結論部ごとに、ユーザが複数の正しい例や誤った例をマルチモーダルに与えることにより、ルール条件部やルール実後部の記述を自動的に生成すること、(6)さらに、上記のようにユーザが操作の例を明示的に記述しなくても、ユーザの操作履歴をキーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた入力情報や、静止画、動画、音声、効果音等の出力情報、ユーザの操作等からなる操作履歴を保存し、それらの中からよく現われるパターンを検出し、マルチモーダルルールを自動生成すること、の手段を備えた構成としてある。

【0008】

【発明の実施の形態】以下、図面を用いて本発明の実施例を説明する。

【0009】図1は本発明の第一の実施例を示す。

【0010】本実施例は、図1に示すように、モニタ1、スピーカ3、カメラ4、マイク5などから構成されるパーソナルコンピュータに存在するエージェント2が対象となる。エージェント2は、このパーソナルコンピュータの前のユーザのキーボード、マウス、ペン、音声、身振りなど複数のメディアを用いた指示に基づき、ユーザに代わってパーソナルコンピュータシステム自身

やこのパーソナルコンピュータ上で動作しているアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答する機能を有している。

【0011】このエージェントソフトウェアの動作内容をユーザがプログラミングすることなく、自由自在に追加、訂正するための機能、すなわち、知的エージェント構築支援機能を有することが本システムの特徴であり、そのメインウィンドウ7の様子がモニタ6内に表示されている。メインウィンドウ7は、エージェントの動作を規定するためのマルチモーダルルールを表示しており、マルチモーダルルール条件部8とマルチモーダルルール実行部9から構成されている。

【0012】図2以降を用いて、知的エージェント構築支援機能について説明する。エージェント2をマウスでクリックすると、そのエージェントに関するメニュー21が現われる。メニュー21には、マルチモーダルルール作成22、操作履歴保存開始23、操作履歴からの学習24の3つのサブメニューが表示されている。

【0013】図3を用いて全体の動作フローを説明する。まず、モニタ1上に表示されている知的エージェント2上をマウスでクリックする動作によるメニュー起動操作があったかどうか判定し(31)、もしメニュー起動動作があれば、図2に示したメニューを表示する(32)。

【0014】ユーザがメニュー中の“マルチモーダルルール作成”を選択した場合(33)、図5に示すマルチモーダルルールエディタを起動する(34)。マルチモーダルルールエディタを用いて、エージェントの動作を規定するマルチモーダルルールを作成する方法については、図5を用いて後述する。

【0015】ユーザがメニュー中の“操作履歴保存開始”を選択した場合(35)、履歴記憶スイッチをオンに設定する(36)。

【0016】ユーザがメニュー中の“操作履歴からの学習”を選択した場合(37)、履歴ファイルを解析し、マルチモーダルルールを自動生成する(38)。

【0017】さらに、履歴記憶スイッチがオンである場合(39)、マルチモーダルルールの自動生成に必要なマルチモーダル履歴記憶を履歴ファイルに書き出す(40)。マルチモーダルルールの自動生成に必要なマルチモーダル履歴情報に関しては、図9を用いて後述する。

【0018】以上のように、ユーザがメニュー中の“マルチモーダルルール作成”を選択した場合、マルチモーダルルールエディタが起動され、図5に示すウィンドウが表示される。このウィンドウを操作することにより、マルチモーダルルールのユーザによる作成やユーザの操作例を基にした自動作成を行うことができる。

【0019】そこで、まず、図5に示すウィンドウが現われたときの動作を図4に示す動作フローを用いて説明

10

20

30

40

50

する。マルチモーダルルールエディタのウィンドウ51には、ルールの名称52とルールの定義情報であるルール条件部55とルール結論部58が表示されている。学習開始ボタン53と学習終了ボタン54は1つのルール全体の学習をコントロールするボタンであり、学習開始ボタン56と学習終了ボタン57はルールの条件部の学習をコントロールするボタン、学習開始ボタン59と学習終了ボタン60はルールの実行部の学習をコントロールするボタンである。

【0020】このマルチモーダルルールエディタのウィンドウ51が表示されている状態において、ルール学習ボタン53が押された場合(41)、学習終了ボタン54が押されるまで図9に例示した形式で操作履歴を記憶し、結果を解析し1つのルールを生成する(42)。また、ルール条件部学習ボタン56が押された場合(43)、学習終了ボタン57が押されるまで図9に例示した形式で操作履歴を記憶し、結果を解析し1つのルール条件部を生成する(44)。また、ルール実行部学習ボタン59が押された場合(45)、学習終了ボタン60が押されるまで図9に例示した形式で操作履歴を記憶し、結果を解析し1つのルール実行部を生成する(46)。操作履歴の保持形式や学習方法は図9を用いて後述することにして、以下、マルチモーダルルールをユーザ自身で定義する方法47について説明する。

【0021】図6は、図5のルール条件部55を拡大したものである。入力メディアとして、「音」61、「音声」62、「キーボード」63、「マウス」64、「ペン」65、「身振り」66、「画面」67の計7種が用意されている。「音」と「音声」の違いは、「音」61は言葉にならない効果音を表わすのに対し、「音声」62は言語化可能な音を表わす。また、「画面」67はその時パーソナルコンピュータのモニタ上に表示されている画面の一部分を表わす。これについては図8を用いて後述する。

【0022】7種の入力メディアごとに、それぞれエディタウィンドウ68が用意されている。メディアごとにエディタウィンドウの動作は異なるが、時間座標69を持つ。また、開始時刻を明示するために、三角印70を用いる。

【0023】「音」用のエディタウィンドウでは、音が発生されるべき時刻を選択し、音を実際に発生させてマイクで収集し、その波形が表示される。

【0024】「音声」用のエディタウィンドウでは、音声が発生されるべき時刻を選択し、その音声のテキストを入力する。開始時刻は三角印70で明示してもよい。

【0025】「キーボード」用のエディタウィンドウでは、キーボードから入力されるべきテキストを入力する。

【0026】「マウス」用のエディタウィンドウでは、マウスから入力されるべきクリックを座標と時刻を表わ

す三角印で入力する。2つの連なった三角印はダブルクリックを表わす。

【0027】「ペン」用のエディタウィンドウでは、ペンから入力されるべきストロークデータを入力する。

【0028】「身振り」用のエディタウィンドウでは、身振りが入力されるべき時刻を選択し、身振りを実際に発生させてカメラで収集し、その様子が表示される。

【0029】「画面」用のエディタウィンドウでは、パーソナルコンピュータの画面で表示されるべき画面の一部分を画面から切り取り入力する。

【0030】以上のように、各入力メディアごとに記述することにより、マルチモーダルルールの条件部を定義することができる。この条件部に規定されているメディアで記述されている入力が、指定された時刻にそれぞれ観測された場合、このマルチモーダルルールは実行可能であると判定される。

【0031】次に図7を用いてマルチモーダルルールの実行部の定義方法を説明する。図7は、図5のルール実行部58を拡大したものである。出力メディアとして、

「音」71、「音声」72、「テキスト表示」73、「アニメーション」74、「コマンド実行」75、「画面」76の計6種が用意されている。「音」と「音声」の違いは、入力部と同様、「音」は言葉にならない効果音を表わすのに対し、「音声」は言語化可能な音を表わす。

【0032】6種の出力メディアごとに、それぞれエディタウィンドウ79が用意されている。メディアごとにエディタウィンドウの動作は異なるが、時間座標77を持つ。また、開始時刻を明示するために三角印78を用いることなどは、入力部と同様の意味を持つ。

【0033】「音」用のエディタウィンドウでは、音が発生されるべき時刻を選択し、音を実際に発生させてマイクで収集し、その波形が表示される。

【0034】「音声」用のエディタウィンドウでは、音声が発生されるべき時刻を選択し、その音声のテキストを入力する。開始時刻は三角印78で明示してもよい。

【0035】「テキスト表示」用のエディタウィンドウでは、モニタに文字列として表示すべきテキストを入力する。

【0036】「アニメーション」用のエディタウィンドウでは、モニタに動画として表示すべきアニメーションを入力する。

【0037】「コマンド実行」用のエディタウィンドウでは、パーソナルコンピュータで実行すべきコマンドを明示的に記述する。図7の例では、あるファイル(File01)を削除するコマンド>Delete File01が記述されている。

【0038】「画面」用のエディタウィンドウでは、本ルールを実行後、パーソナルコンピュータの画面で表示されるべき画面の一部分を画面から切り取り入力する。

10

20

30

40

50

この解釈に関しては図8を用いて後述する。

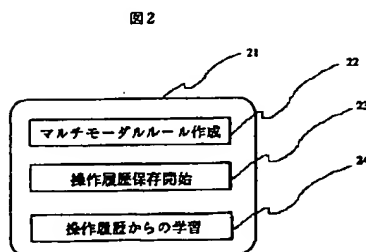
【0039】以上のように、各出力メディアごとに記述することにより、マルチモーダルルールの実行部を定義することができる。

【0040】マルチモーダルルールを実行するとルールの実行部に規定されたメディアで、そこに記述された内容を、指定された時刻に出力することになる。例外は、「コマンド実行」75と「画面」76である。「コマンド実行」75に記述された内容は、指定された時刻にそのコマンド自身が実行され、「画面」76の場合は、指定された時刻に、そこに記述された画面の状態になるようなコマンド列が推定され、順々に実行されることになる。

【0041】図8に、「画面」に関する定義のみからなるマルチモーダルルールの例を示す。条件部にある画面の記述81はファックス受信時にパーソナルコンピュータのデスクトップに現われるアイコンの変化後の状態を表わしている。すなわち、パーソナルコンピュータにファックスが到着するとこのアイコン81が現われる。ルール実後部の記述82はそのファイルをホルダAにのせる、つまり、格納する操作を行う場合の画面の状態を表わしている。従って、本マルチモーダルルールはファックス受信時に実行可能となり、受信したファイルがホルダAに保存されるようなコマンドシーケンスが実行されることになる。

【0042】図9にルール全体の学習、ルール条件部の学習、ルール結論部の学習に用いられる履歴情報の格納形式を示す。格納形式91は、ルール条件部、結論部の定義形式を融合したものであり、項目としては、「音」92、「音声」93、「キーボード」94、「マウス」95、「ペン」96、「身振り」97、「画面」98である。これらからルール条件部、ルール結論部を直接生成することが可能である。また、共通操作の抽出は従来知られているクラスタリング手法を用いる。

【図2】



【0043】

【発明の効果】以上説明したように本発明の実施例によれば、キーボード、マウス、ペン、音声、身振りなど複数のメディアを用いたユーザの指示に基づき、ユーザに代わってコンピュータやアプリケーションソフトウェア等を制御し、静止画、動画、音声、効果音等の複数のメディアを用いて応答する知的なエージェントを、ユーザ自身で、プログラミングを行うことなく定義することが可能となる。

10 【図面の簡単な説明】

【図1】本発明の第一の実施形態を示す図である。

【図2】エージェントに関するメニューを示す図である。

【図3】全体の動作フローを示す説明図である。

【図4】マルチモーダルルールウィンドウの動作フローを説明する図である。

【図5】マルチモーダルルールエディタの全体を表わす図である。

20 【図6】マルチモーダルルールエディタの条件部を説明する図である。

【図7】マルチモーダルルールエディタの実後部を説明する図である。

【図8】画面に関する記述のみからなるマルチモーダルルールの例を示す図である。

【図9】マルチモーダル履歴情報の例を示す図である。

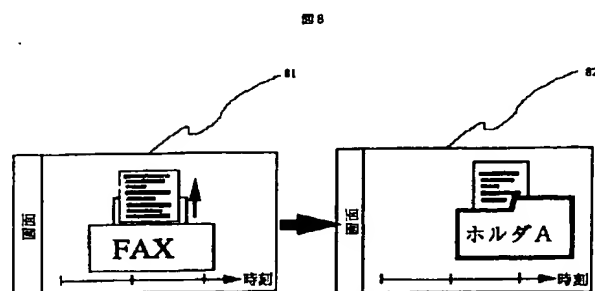
【符号の説明】

1…パーソナルコンピュータのモニタ、2…エージェント、3…スピーカ、4…カメラ、

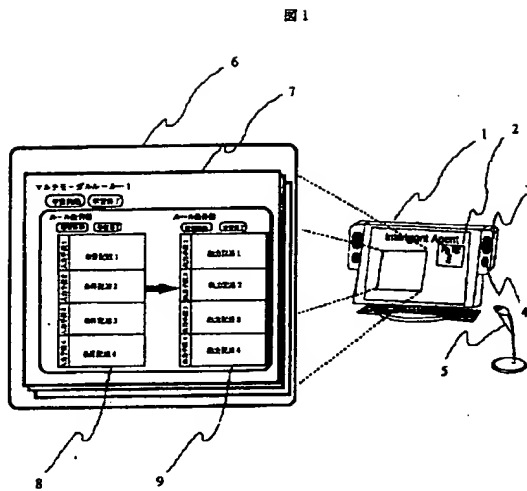
5…マイク、7、51…マルチモーダルルールウィンドウ、8、55…マルチモーダルルール条件部、

9、58…マルチモーダルルール実行部、21…エージェントメニュー、81、82…画面のみの定義からなるマルチモーダルルールの例、91…マルチモーダル履歴記憶情報。

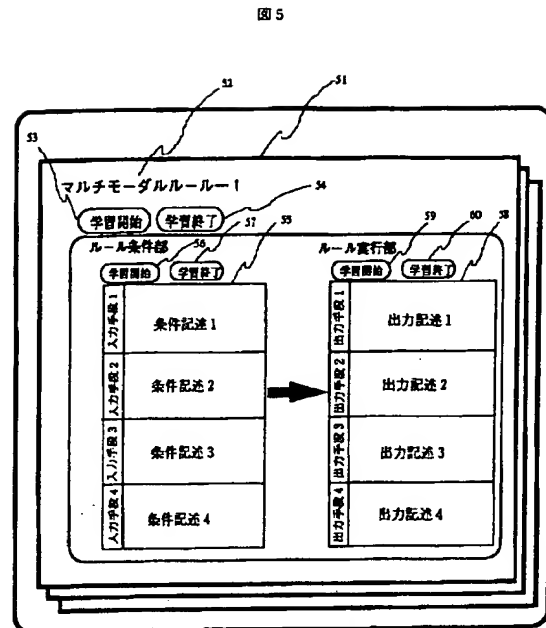
【図8】



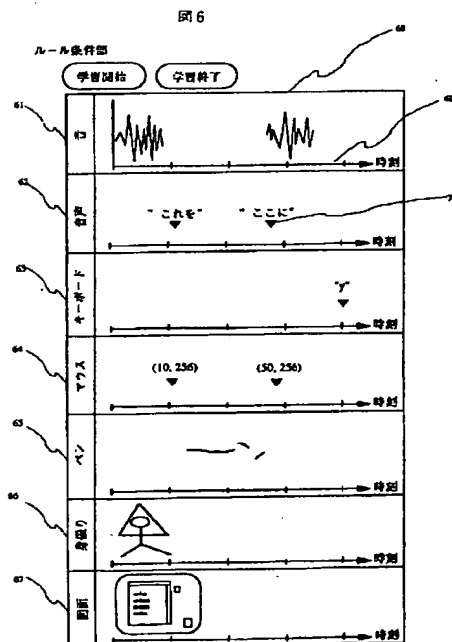
【図1】



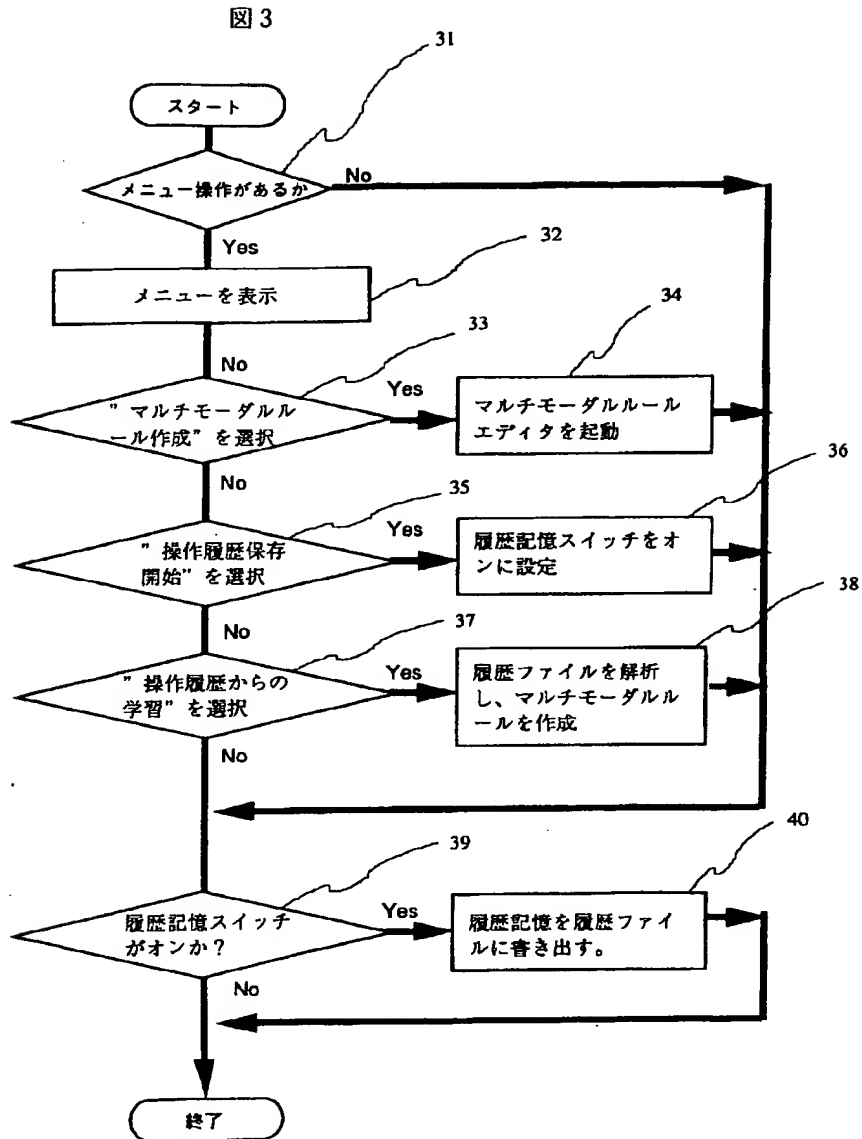
【図5】



【図6】



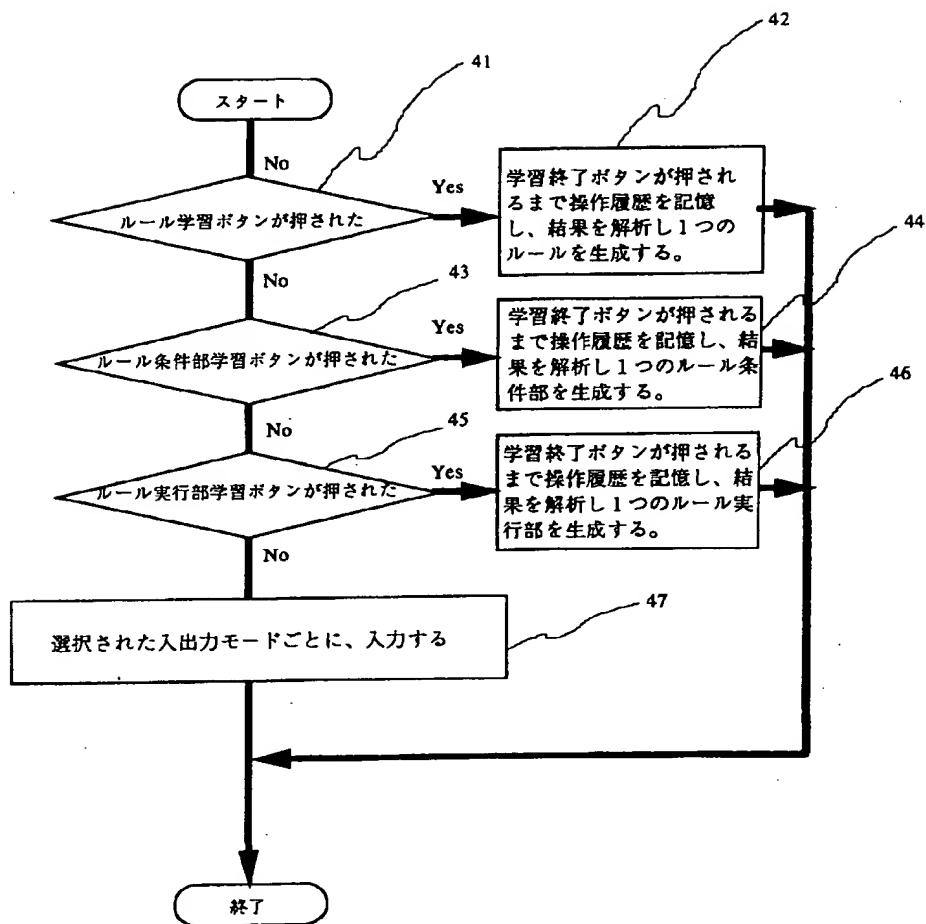
【図3】



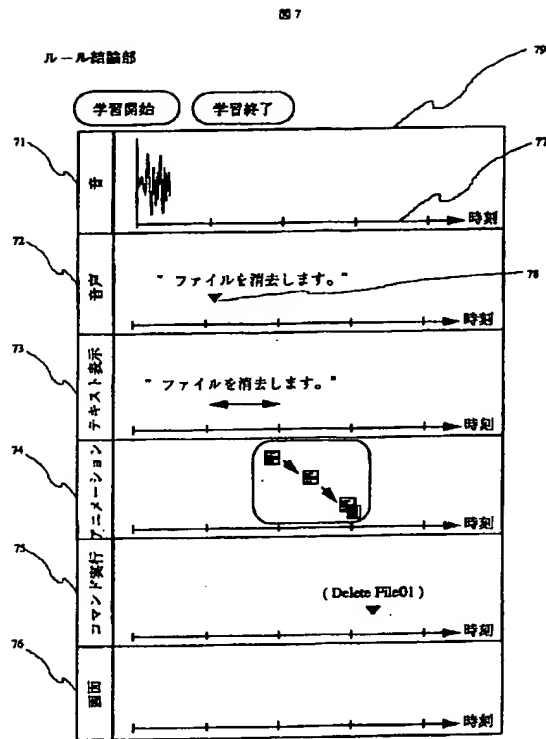


【図4】

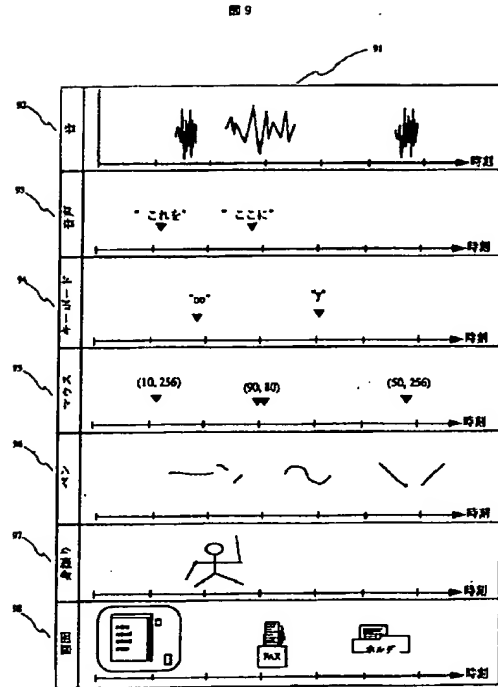
図4



【図7】



【図9】



フロントページの続き

(72)発明者 富田 民則  
神奈川県川崎市麻生区王禅寺1099番地株式  
会社日立製作所システム開発研究所内

(72)発明者 青島 弘和  
神奈川県川崎市麻生区王禅寺1099番地株式  
会社日立製作所システム開発研究所内